多维度视角下国外科学数据管理的研究脉络分析*

李煜[1,2] 刘虹[1] 孙建军[1]
[1] 南京大学信息管理学院 南京 210023 [2] 西北大学图书馆 西安 710127

摘要:[目的/意义]大数据环境下,科学数据管理研究具有重要的理论和现实意义。对科学数据管理研究脉络进行可视化分析,可以为我国该主题研究提供参考与借鉴。[方法/过程]基于WoS和Google Scholar,利用Citespace、Ucinet、TDA等分析工具,探析国际学术界科学数据管理研究发展的演进过程,全面揭示科学数据管理研究的历史、现状和发展方向。[结果/结论]从时间分布图谱、空间分布图谱、主题演进图谱和知识基础图谱4个维度反映科学数据管理的研究现状,并结合国内相关研究及发展状况提出相关建议。

关键词:科学数据 数据管理 知识图谱 可视化分析
分类号:G250
DOI:10.13266/j.issn.0252-3116.2018.13.014

1 引言

随着信息技术不断深入地发展,科学研究已经全面进入到"第四范式",各个学科的科学数据日益丰富并且迅速累积,共同构成了科研活动的大数据环境。科学数据作为具有重要科研价值、社会价值及经济价值的战略性资源,其管理和利用已经被国内外各级政府和科研机构高度重视。2002年,我国科技部开始实施"国家科学数据共享工程"[1],迄今为止已经在资源环境、农业、人口与健康等多个学科领域开展了科学数据管理与共享工作,为我国科学数据管理工作的深入开展奠定了良好基础。2015年,国务院印发《促进大数据发展行动纲要》[2],从顶层设计的角度提出了我国实施大数据发展的战略部署,是指导我国科学数据管理与利用的纲领性文件。

一般认为,科学数据是指人类在认识世界和改造世界的科技活动中通过实验、观测、调查等方式所产生的基础性原始数据,以及按照特定的需求加工产生的数据集和相关信息[3]。在大数据背景下,科学数据的范畴发生了很大的变化,广义的科学数据不仅包括自然科学领域的数据,同时也包括人文社会科学领域研究中产生的各类数据,甚至还包括科研工作者的个人数据、互联网上的各种数据等。科学数据具有学科领域分布广泛、数据结构多样、数据共享困难等特点,因此科学数据管理的研究和实践活动对整个社会、科研机构或者科研人员来说都具有重要的意义。

近年来,国内外学术界从不同角度对科学数据管理进行了研究,其关注点主要集中在以下两个方面:①对美国、英国、澳大利亚、加拿大等开展科学数据管理工作较早国家的成功经验进行分析与借鉴。刘桂锋等选取斯坦福大学、哥伦比亚大学和华盛顿大学3所高校作为样本,对美国高校中图书馆参与科学数据管理的实践进行了分析与探讨[4];司莉和辛娟娟利用内容分析法对英美20所大学颁布的科学数据管理与共享政策进行了比较和分析,并对我国制定相关政策提出了建议[5]。②对基础设施、服务需求、资金来源等科学数据管理的业务实践等方面进行调研与分析。唐燕花通过案例分析法探究高校开展科学数据管理服务的主要构成要素,并根据这些实践要素和国内高校开展科学数据管理服务的现状提出有效建议[6];卫军朝和张春芳选取国内外10余个科学数据管理平台,通过对建设现状、建设目标、数据来源、经费情况等多个维度进行横向对比,探究了不同平台之间的差距,发现了我国科学数据管理实践中值得借鉴和参考的经验[7]。纵观已有的研究成果可以发现,目前学术界对科学数据管理研究主要为定性研究,大部分侧重于对国外单个或

* 本文系国家自然科学基金面上项目"融合范式视角下的链接分析理论集成框架及实证研究"(项目编号:71273125)研究成果之一。
作者简介:李煜(0000-0001-8385-9862),馆员,博士研究生,E-mail:leeyuk@nwu.edu.cn;刘虹(0000-0003-4943-7855),博士研究生;孙建军(ORCID:0000-0003-2144-8253),院长,教授,博士。
收稿日期:2018-01-10 修回日期:2018-03-28 本文起止页码:111-118 本文责任编辑:徐健

多个机构开展科学数据管理实践的调查与分析，无法从全局范围内展示国内外科学数据管理的研究现状。因此，笔者采用科学计量学方法，利用 Citespace、Ucinet、TDA 等多种数据分析和可视化工具，更加系统地从数据维度反映世界各国在科学数据管理领域的研究现状，以期为我国科学数据管理研究和实践提供参考与借鉴。

2　科学数据管理可视化方法体系设计及实现

2.1　可视化方法体系框架

为了更加全面地分析科学数据管理领域的研究现状，笔者利用图 1 的方法体系框架从多个角度和维度直观展示该领域研究的主体特征、热点分布及变化趋势。

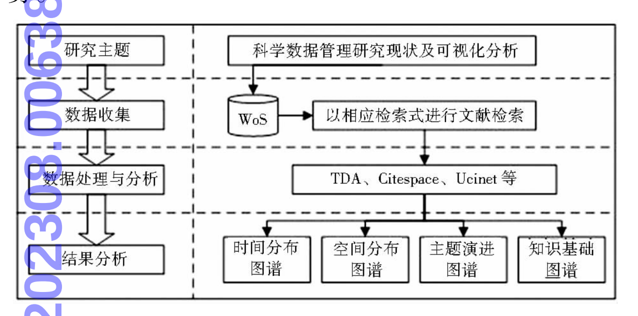


图 1　科学数据管理的可视化分析方法体系框架

首先，根据研究主题，构建"科学数据管理"的检索式。为了尽可能全面地展示科学数据管理的研究现状，选用国际核心学术期刊数据库 Web of Science 中的 SCI、SSCI 和 A&HCI 作为数据来源库，利用"scientific data management""scientific data curation""research data curation"等相关检索词作为主题词进行检索，检索时间截至 2017 年 11 月 20 日，排除误检后共获得科学数据管理研究论文 336 篇。

其次，利用 Thomson Data Analyzer 等数据分析工具对 336 篇研究论文进行规范化处理，包括剔除无效关键词、合并相似关键词、构建关键词共现矩阵、规范研究机构名称、合并同一机构的不同表现形式等。

再者，对科学数据管理领域的知识产出进行可视化展示与分析，包括时间分布图谱、空间分布图谱、主题演进图谱及知识基础图谱 4 个维度。

2.2　可视化方法实现机制

以规范化处理后的 336 篇研究论文及其引文数据为基础，从以下 4 个方面实现科学数据管理研究的可视化展示与分析：

首先，绘制科学数据管理研究的时间分布图谱，从时间序列维度展示该主题的研究足迹，分析其知识产出的发展趋势。

其次，绘制科学数据管理的空间分布图谱，从国家和机构两个层面展示该主题科研成果在空间上的分布情况，识别不同国家、不同科研机构竞争优势。

再者，绘制科学数据管理的主题演进知识图谱。利用 Ucinet 对高频关键词进行共现分析，展示科学数据管理研究的知识网络；利用 Citespace 绘制关键词时区视图，把握研究主题随时间的变化趋势；利用内容分析法对研究热点进行深入剖析，总结归纳科学数据管理研究的主题分布。

最后，绘制科学数据管理的知识基础图谱。利用 Citespace 对该主题研究的引文共被引情况进行聚类，通过知识基础图谱展示科学数据管理主题的理论基础；并利用 Google Scholar 对该主题研究的高被引参考文献做进一步深入分析，了解科学数据管理研究的核心知识基础。

3　科学数据管理的知识图谱分析

3.1　时间分布图谱

时间分布图谱能够从时间维度展现科学数据管理研究的总体图景，识别该主题研究的发展历程和发展趋势。因此，对 336 篇科学数据管理研究论文进行统计分析，其时间分布图谱如图 2 所示（因检索时间截至 2017 年 11 月 20 日，2017 年论文数量暂不能代表全年趋势）：

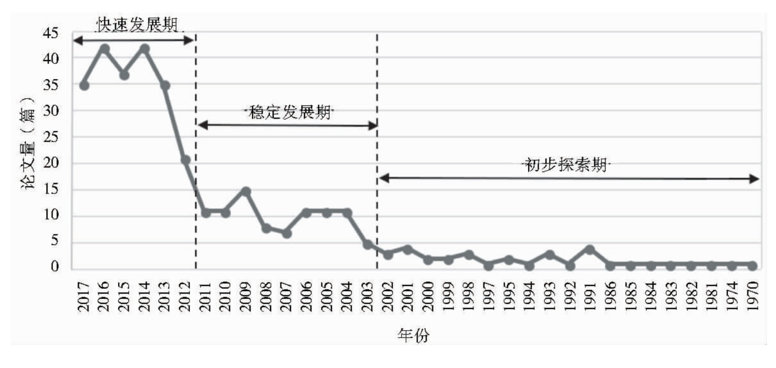


图 2　科学数据管理研究时间分布图谱

从图 2 可以看出，学术界对该主题的关注程度随着时间发展不断提高，知识产出数量呈现出指数增长趋势。科学数据管理的研究足迹按照时间序列可以划分以下 3 个阶段：

李煜，刘虹，孙建军. 多维度视角下国外科学数据管理的研究脉络分析[J]. 图书情报工作,2018,62(13):111－118.

（1）初步探索期（论文数量＜5 篇）：1970－2002 年。在初步探索期，科学数据管理主题的知识产出数量较少，论文数量低于 5 篇。该主题研究最早可以追溯到 1970 年，随后直到 2002 年的 30 多年时间里，在航空航天工程[8]、环境科学[9]、化学[10]、生物医药[11]、计算机科学[12]等多个学科领域都对科学数据管理主题进行了探索性研究。

（2）稳定发展期（5 篇＜论文数量＜15 篇）：2003－2011 年。2003 年 10 月，德国马普学会召开柏林会议并发布《关于自然科学与人文科学知识开放获取的柏林宣言》，包括中国在内的多个国家签署了"柏林宣言"以支持科学研究成果和原始科学数据的开放获取；2004 年 1 月，世界经济合作与发展组织（Organization for Economic Co-operation and Development，OECD）的 34 个成员国签署了科技政策委员会发布的《开放获取公共资助研究数据的宣言》，明确了科学数据开放的范围和定义。随着一系列科学数据管理政策的发布，科学数据管理研究进入稳定发展期，这一阶段的知识产出能力相比初步探索期有了较大幅度的提升。

（3）快速发展期（论文数量＞20 篇）：2012－2017 年。美国国家自然科学基金委员会（National Science Foundation，NSF）于 2010 年提出申请 NSF 资助项目必须提交"数据管理计划"；澳大利亚研究理事会（Australian Research Council，ARC）和国家卫生与医学研究理事会（National Health and Medical Research Council，NHMRC）分别于 2012 年、2013 年对科研资助项目的研究成果提出管理和共享要求；英国研究理事会（Research Councils UK，RCUK）于 2011 年要求受资助项目在不损害知识产权的情况下对科学数据实施共享；2014 年，中国科学院（Chinese Academy of Sciences，CAS）和国家自然科学基金委员会（National Natural Science Foundation of China，NSFC）也分别发布了科研项目成果的开放获取声明。科学基金有利于科研人员创新思想、获取资源、合作交流，对科学研究活动具有巨大的支持作用，能够促进科学研究跨越式的发展。在众多基金政策的影响下，科学数据管理的研究进入快速发展期，科研产出数量呈现出爆炸性增长趋势。

3.2 空间分布图谱

对科学数据管理研究领域的国家和机构进行调查与分析，可以从空间角度识别不同国家和科研机构在该领域的竞争优势，了解各机构在该领域的合作和交流情况，能够为我国学术界研究科学数据管理提供参考和借鉴。如图 3 所示：

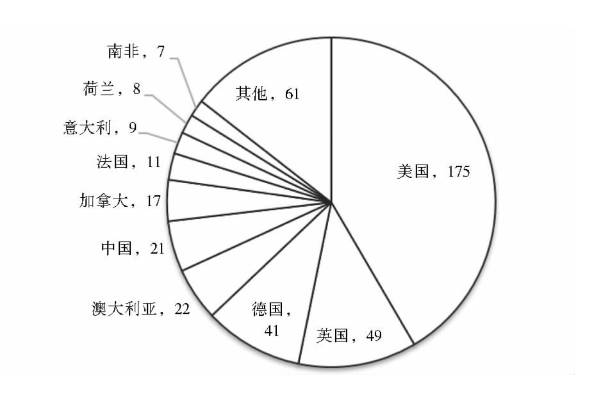


图 3　科学数据管理的国家分布图谱

从国家层面来看，美国在科学数据管理主题领域的知识产出总量遥遥领先，其论文数量占比约为 52.1%，这与美国政府、基金资助机构、科研机构及科研人员长久以来对科学数据政策制定、科学数据资源管理等相关活动高度重视密切相关。除美国外，英国、德国、澳大利亚也在科学数据管理领域发表了大量的研究成果。根据调查，以上 3 个国家在科学数据资源的建设规模、注册或加入世界数据系统（World Data System，WDS）和 DataCite 等科学数据国际重要组织、建设全球性的科学数据仓储系统（如 Re3data. org）[13]等方面均具有较大的优势。同时由图 3 可以看出，我国虽然在科学数据管理领域取得了一定数量的成果，但是与美国、英国、德国等发展较早的科学数据强国相比来说，还存在较大的差距。

为了进一步了解我国与数据强国的科研机构之间存在的发展差异，从知识产出能力和学术影响力两个角度对科学数据管理研究的核心科研团体进行分析。度中心性可以测量机构合作网络中某个机构与其他所有机构相互联系的程度，计算公式如下所示：

$$C_D(N_i) = \sum_{j=1}^{g} x_{ij}(i \neq j) \qquad \text{公式(1)}$$

其中，$C_D(N_i)$表示机构 i 的度中心性，g 表示科研机构总数量，$\sum_{j=1}^{g} x_{ij}$表示机构 i 与其他 g－1 个 j 机构直接进行合作的文章数量。论文数量越多，证明机构在科学数据管理领域的知识产出能力越强；而度中心性越高，证明该机构在科学数据管理领域的学术影响力越大。选取发文量在 5 篇以上的学术机构进行度中心性计算，结果见表 1。

由表 1 可以看出，美国国家航空航天局（NASA）是在科学数据管理领域发表论文数量最多的机构，NASA 最早在 1993 年 3 月就为每个科学研究项目制定了项目数据管理计划（Project Data Management Plan，PDMP）[14]，通过对各级各类科学数据进行管理和存档，为企业、学术界以及公众提供数据产品、数据信息

表 1　科学数据管理领域核心机构分布

序号	学术机构	论文量(篇)	度中心性	国家
1	美国国家航空航天局	12	6.000	美国
2	谢菲尔德大学	11	6.000	英国
3	牛津大学	8	9.000	英国
4	约翰霍普金斯大学	7	10.000	美国
5	哈佛大学	7	4.000	美国
6	中国科学院	7	1.000	中国
7	法国国家科学研究院	7	0.000	法国
8	匹兹堡大学	6	7.000	美国
9	芝加哥大学	6	4.000	美国
10	加州大学戴维斯分校	6	3.000	美国
11	武汉大学	6	0.000	中国
12	南佛罗里达大学	5	0.000	美国
13	伊利诺伊香槟分校	5	6.000	美国
14	哥廷根大学	5	0.000	德国
15	爱丁堡大学	5	6.000	英国
16	美国国家大气研究中心	5	6.000	美国

或者数据使用工具等多种服务，促进科学数据的研究与应用[15]。同时可以看出，美国约翰霍普金斯大学是科学数据管理领域与其他科研机构合作最广泛的高校，约翰霍普金斯大学的科学数据管理服务(Johns Hopkins University Data Management Services，JHUDMS)[16]可以为科研人员和研究机构提供科学数据管理计划咨询、科学数据归档、科学数据管理培训等多种个性化服务，并且与伊利诺伊香槟分校、德克萨斯大学休斯顿健康科学中心、亚利桑那州立大学、美国 BioFortis 公司等多个机构进行了合作研究。

综合图 3 和表 1 可以得出，美国和英国不仅是科学数据管理领域的数据大国，同时也是数据强国。在法律法规、基金资助条件、科研机构数据管理政策等多种内外部因素的共同驱动下，产生了以美国航空航天局、谢菲尔德大学、牛津大学、约翰霍普金斯大学、哈佛大学等一系列著名高校及科研机构为代表的核心科研团体。而我国仅有中国科学院和武汉大学两个机构在科学数据管理领域发表了 5 篇以上国际核心期刊论文，无论是在知识产出能力还是在学术影响力方面都与数据强国存在较大的差距。

3.3　主题演进图谱

3.3.1　关键词共现知识图谱　对科学数据管理研究的高频关键词进行共现分析，可以厘清科学数据管理研究的发展现状。利用 TDA 软件对 336 篇科学数据管理研究论文的关键词进行清洗，并提取核心关键词构建关键词共现矩阵，使用 Ucinet 对核心关键词共现矩阵进行可视化分析，结果见图 4。在关键词共现知识图谱中，节点的大小用度中心性表示，节点越大代表该关键词在科学数据管理研究网络中的影响越大，是该领域的研究热点。

从图 4 可以看出，数据管理(data management)、研究数据管理(research data management)、数据共享(data sharing)、数据监管(data curation)4 个关键词居于整个知识网络的中心位置，是科学数据管理研究中最核心的热点问题。其中，数据管理作为词频及中心性最高的关键词，其研究范围非常广泛，研究内容贯穿数据共享、数据处理、数据存储、数据获取、数据分析等科学数据生命活动全周期，并且拓展到系统生物学、信息管

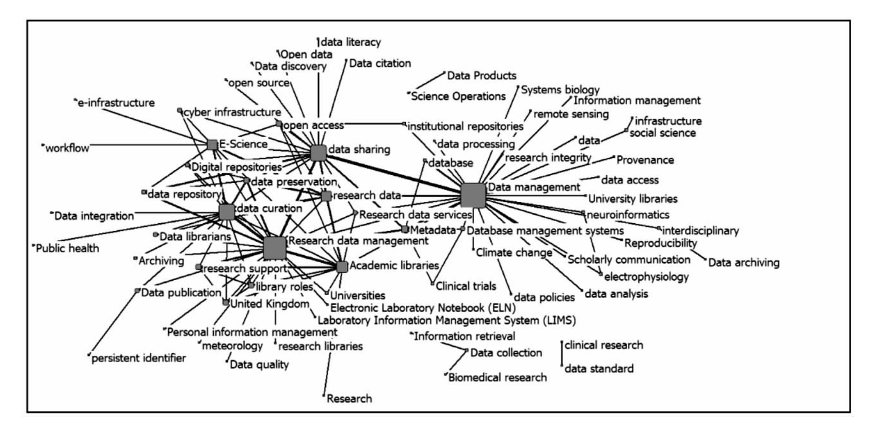


图 4　科学数据管理核心关键词共现知识图谱

理、遥感、社会科学、神经信息学、电生理学、跨学科等全学科领域；研究数据管理作为科学数据管理的子范畴，其研究内容除了数据保存、数据监管、数据共享等普遍适用的数据管理技术与方法外，更多地从学术图书馆、大学、图书馆角色、研究支持、个人信息管理、实验室信息管理系统等科研活动支撑角度进行研究；科学数据共享是促进学术交流、提高科学数据利用效率的重要手段，数据共享也是科学数据管理研究中词频和中心性较高的关键词，数据共享与开放获取、数据发现、开放数据、数据素养、数据引用、数据保存等关键词紧密关联；数据监管是 E-science 环境下科学数据长期保存、管理和利用的新兴研究热点，因此数据监管与数据保存、数字存储、数据整合、存档、数据仓库、E-science 等关键词产生了高频共现，为科学数据管理研究与应用提供了数据保障。

李煜，刘虹，孙建军. 多维度视角下国外科学数据管理的研究脉络分析[J]. 图书情报工作,2018,62(13):111 - 118.

3.3.2　研究前沿时区知识图谱　高频关键词时区图谱可以从时间维度展示科学数据管理研究热点演进和变化趋势。利用 Citespace 绘制核心关键词时区图谱，结果如图 5 所示：

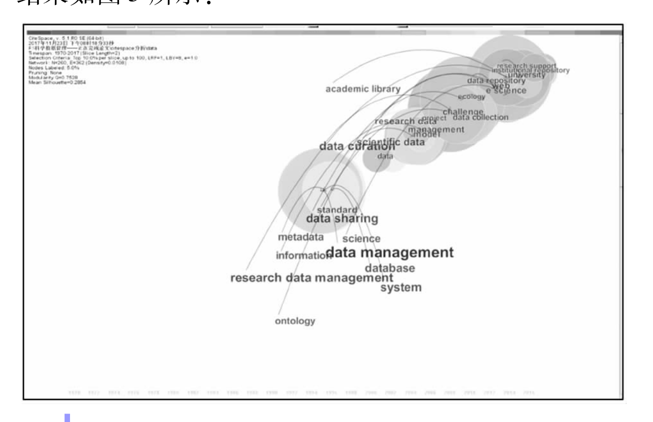


图 5　科学数据管理核心关键词时区知识图谱

由图 5 可以看出，在科学数据管理研究的 3 个发展阶段中，科研人员对该主题呈现出不同的关注点：

在初步探索期(1970 - 2002 年)，科学数据管理研究知识产出数量较少，研究内容集中于以信息技术为载体的科学数据管理系统的开发与利用，例如对数据库(database)、系统(system)、元数据(metadata)等问题的研究。该阶段研究虽然处于初步发展期，但却在研究范围上得到充分扩展，例如环境科学、航空航天工程、化学、生物医药等多个自然科学领域都在科学数据管理中做出了探索性研究。

进入稳定发展期(2003 - 2011 年)后，科学数据管理研究论文数量出现了大幅度增长，研究内容进一步向纵深方向发展，此阶段的科学数据管理研究突破数据管理系统的局限，开始向数据共享(data sharing)、数据监管(data curation)、数据收集(data collection)等涉及科学数据管理生命全周期的方向发展；研究范围也开始突破自然科学领域范畴，人文社会科学领域以及跨学科领域的科学数据管理开始受到研究人员的关注。

在快速发展时期(2012 - 2017 年)，科学数据管理在前两个阶段的基础上，不仅沿着科学研究维度纵深发展，并且也向着科研支撑维度横向发展。例如数据共享(data sharing)、数据仓库(data repository)依然是科学数据管理研究的重点内容，同时学术图书馆(academic library)、机构知识库(institutional repository)、科研支持(research support)等与科学数据管理服务相关的主题也成为该阶段研究的重点内容。

3.3.3　研究热点主题分布　核心关键词共现知识图谱和时区知识图谱在一定程度上展示了科学数据管理领域热点的分布和演进情况。为了更加深入和全面地了解研究热点主题分布，通过对科学数据管理研究的主题词、学科分布等信息进行定量与定性综合分析，将该领域研究分为 4 个方面。下文采用内容分析法对科学数据管理领域的 4 大研究主题进行具体的论述：

(1)基于管理主体的科学数据管理研究。对 336 篇研究论文的题名、关键词词频、摘要等进行内容分析可以看出，对各级管理主体的科学数据管理活动进行分析是该领域最为集中的研究内容之一。基于管理主体的科学数据管理研究包括跨国组织、国家、国内联盟及学术机构 4 个层面。J. Wei 等人调查了在大数据环境下，全世界最大的社会科学数据中心 ICPSR 如何存储、监管和共享社会科学数据[17]；J. Kim 利用问卷法调查了 190 名韩国教授对科学数据管理与共享的看法，以期为韩国制定数据政策及数据管理实践提供参考与借鉴[18]；J. Dierkes 等介绍了哥廷根 E-Research 联盟向其成员机构提供基于科学数据生命周期的数据管理创新服务[19]；M. Witt 介绍了普渡大学如何收集和管理科学数据，从而更好地为多个学科的科研活动提供支持服务[20]。

(2)基于学科的科学数据管理研究。根据科学数据的定义可以看出，早期的科学数据主要产生在自然科学领域，对科学数据管理的研究也集中在生物学、医学、气象学、地球科学等学科。随着大数据技术的发展，科学数据的范畴有了很大变化，科学数据管理研究扩展到了语言学、人口学、心理学等人文社会科学领域及跨学科领域。S. Adcock 等人分析了地球化学数据管理中存在的问题及面临的挑战，并介绍了加拿大地质调查局在地球化学数据管理中采用的方法和获得的经验[21]；P. Fankhauser 等人以日耳曼语言学为例研究了人文科学研究中的科学数据管理问题[22]；R. Downs 分析了目前不同学科领域科学数据管理存在的不足，提出了新的模型来管理跨学科的科学数据[23]。

(3)基于技术的科学数据管理研究。在数据密集型科学研究的背景下，科学数据也呈现出数据量爆发性增长、数据结构异构并且多样、数据价值密度低等大数据特征。利用各种信息技术对科学数据进行有效管理，实现科学数据的保存、共享和利用成为科学数据管理的重要研究内容。基于技术角度的科学数据管理研究涉及数据采集、数据传输、数据存储、数据处理、数据分析、数据检索、数据可视化等整个科学数据生命周期，包括硬件设施建设、软件工具开发和数据管理平台

等多个技术维度。

（4）基于图书馆视角的科学数据管理研究。高等院校是科学研究的主体机构之一，在各个学科领域都会产生数量庞大的科学数据资源。在基金资助机构的科学数据管理政策、科研人员科学数据存储需求等内外部因素的驱动下，高校图书馆开展科学数据管理服务势在必行。在大数据环境下，科学数据管理服务为图书馆的发展带来了新的机遇和挑战，国际学术界对于科学数据管理活动中图书馆的角色定位开展了诸多讨论。具体包括：制定科学数据管理政策、研究科研人员数据管理与共享意愿、提供基于科学数据生命周期的管理服务、科学数据管理教育等多个方面，为我国图书馆开展科学数据管理提供了参考。

3.4　知识基础分析

科学数据管理研究自 1970 年出现开始，不断向各个学科扩散，目前已经在自然科学、人文社会科学以及跨学科领域引起了广泛关注。科学数据管理的知识基础由该主题所有文献的引文合集组成，通过文献共被引分析，可以从引用轨迹角度揭示科学数据管理的奠基性知识来源。利用 Citespace 绘制科学数据管理研究文献共被引图谱，结果如图 6 所示：

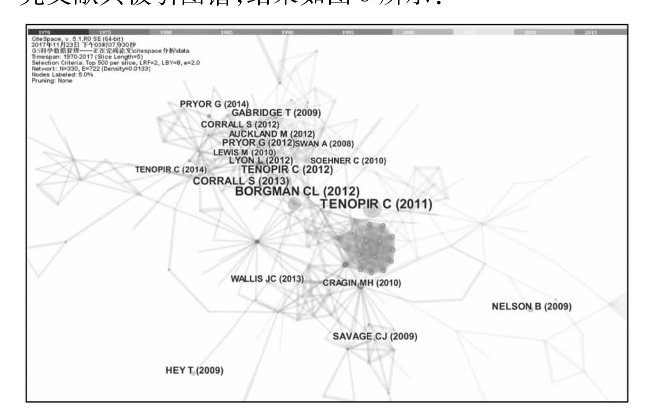


图 6　科学数据管理奠基性知识基础

在知识基础图谱中，每个节点代表一篇引文，节点标签按引文的被引频次标注，即字体越大，该引文的被引频次越高。为了深入地了解对科学数据管理研究具有关键作用的文献信息，选取 Google Scholar 数据库对关键节点进行进一步检索与分析，排名前 10 的共被引文献信息如表 2 所示：

表 2　科学数据管理高被引重要参考文献

序号	作者	题名	文献来源	出版年	被引频次	GS 被引频次
1	C. Tenopir	*Data sharing by scientists: practices and perceptions*	Plos One	2011	29	565
2	C. L. Borgman	*The Conundrum of Sharing Research Data*	JASIT	2012	23	466
3	C. Tenopir	*Academic libraries and research data services: Current practices and plans for the future*	ACRL 白皮书	2012	15	19
4	S. Corrall	*Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research*	Library Trends	2013	15	149
5	T. Hey	*The fourth paradigm: data-intensive scientific discovery*	图书	2009	11	2 105
6	L. Lyon	*The informatics transform: Re-engineering libraries for the data decade*	International Journal of Digital Curation	2012	11	76
7	B. Nelson	*Data sharing: Empty archives*	Nature	2009	10	236
8	J. C. Wallis	*If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology*	Plos One	2013	10	150
9	G. Pryor	*Managing Research Data*	图书	2012	10	78
10	C. J. Savage	*Empirical study of data sharing by authors publishing in PLoS journals*	Plos One	2009	10	208

结合图 6 和表 2 可以看出，C. Tenopir 在 2011 年发表的 *Data sharing by scientists: practices and perceptions* 在科学数据管理领域被引频次最高，该研究选取 1 329 位科研人员调查其科学数据共享的现状，详细分析了科学数据共享中存在的时间不足、缺乏资金、缺少机构支持等障碍和困难，论述了基金资助机构、学科领域、科研人员年龄、工作地区等差异对科学数据管理造成的影响，提出从国家政策及基金层面对科学数据管理进行引导的必要性。该论文在科学数据管理领域被引用 29 次，全球被引 565 次，在其他主题领域也引起了广泛的关注。同时，C. Tenopir 在 2012 年发表的 *Academic libraries and research data services: Current practices and plans for the future* 从科学数据管理服务的角度出发，调查了图书馆在数据密集型科研环境下为科研人员提供科学数据服务的当前现状和未来计划，该研究内容也被科学数据管理领域广泛引用。另外，T. Hey 于 2009 年出版的 *The fourth paradigm: data-intensive scientific discovery* 一书阐释了数据爆炸给数据获取、数

据监管、数据分析及数据共享带来的挑战，讨论了云计算、协作服务和知识库在第四范式环境下的作用。该书全球被引 2 105 次，在世界范围内具有重要影响，同时也是科学数据管理领域的重要参考文献之一。

综上所述，科学数据管理研究的知识基础包括 3 个方面：①在大数据环境下，科学数据爆炸式的增长促使科学研究进入到第四范式，数据密集型科学研究环境为科学数据的管理研究创造了外部条件；②基于避免科学数据丢失、验证科研成果可信度、提高科研成果的影响力、寻求合作机会等一系列目的，科研人员对科学数据管理与共享的需求不断提高，为科学数据管理研究带来了内部驱动；③高校图书馆、学术图书馆等文献情报机构积极参与科学数据管理服务活动，也不断地促进了科学数据管理研究的理论和实践的发展。

4 结论

在数据密集型科研环境下，科学数据呈现出数据量爆发性增长、数据结构异构并且多样性、数据价值密度低等大数据特征，科学数据管理研究与实践对于数据本身、科研人员、科研机构以及全社会都具有重要的意义。笔者采用 WOS 数据库和 Google Scholar 作为数据来源，利用 Citespace、Ucinet、TDA 等多种数据分析工具，从时间分布图谱、空间分布图谱、主题演进图谱及知识基础图谱 4 个维度对国外科学数据管理研究现状进行了全面分析，研究结论如下：

（1）从时间维度看，目前科学数据管理研究正处于快速发展期，在美国国家自然科学基金委员会（NSF）、澳大利亚研究理事会（ARC）、英国研究理事会（RCUK）、英国惠康基金会（Welcome Trust）等基金资助机构的数据管理政策要求下，科研成果增长迅速。国内目前少有基金从政策及制度角度要求被资助者提交“数据管理计划”，在科学数据管理领域还存在广阔的发展空间。

（2）从空间维度看，美国、英国、德国、澳大利亚等开展科学数据研究较早的国家在该领域具有较强的竞争优势，具体包括美国国家航空航天局、谢菲尔德大学、牛津大学、约翰霍普金斯大学等科研机构和高等院校。这些数据强国不仅从国家层面，也从机构层面高度重视科学数据管理研究与实践。

（3）从主题演进的维度看，科学数据管理与共享是贯穿该主题领域的核心研究内容。随着研究的不断深入，研究内容从数据管理发展到科学数据生命全周期，并且在快速发展期受到图书情报学领域的密切关注，学术图书馆、机构知识库、科研支持等科学数据管理服务是当前科学数据管理研究的核心研究主题。

（4）从知识基础维度看，科学数据管理研究的发展建立在数据密集型科研环境、科研人员数据共享与管理需求、图书情报机构积极参与科学数据管理服务三大基础之上。目前科学研究已经进入第四范式，我国在实践领域只有复旦大学、武汉大学、北京大学、中国科学院等少数机构创建了数据服务平台，开展了科学数据管理服务。

我国数据管理研究起步较晚，虽然出现了中国科学院和武汉大学等一些实力较强的科研团队，但是与数据强国之间还存在很大的差距。国内目前对于该主题研究主要集中于对国外机构开展科学数据管理活动的先进经验分析与借鉴方面，研究内容相对单一，有待于从多个主体层面、多个学科范围、多个技术角度进行扩展。具体而言，应该重视以下 4 个方面内容：①从国家、基金、科研机构、出版物多个层面制定完善的数据管理政策，为科学数据管理研究与应用提供强有力的政策支撑；②与美国、英国、澳大利亚等具有丰富经验的科研机构和学术团体建立合作关系，可以提升我国在科学数据管理领域的知识产出能力，同时也能增加我国在国际社会的学术影响力；③加强人文社会科学领域对科学数据的管理与利用，在积极探讨科学数据管理与共享技术、方法的基础上，重视其在全学科领域科学研究及科研服务中的意义；④重视图书情报学领域在科学数据管理活动中的重要地位，学术图书馆、高校图书馆及高校信息管理院系应该丰富科学数据管理研究内容、创新科学数据管理服务、重视科学数据管理人才培养，促进国内科学数据管理的国际化进程。

本文还存在一些研究局限：①选用 Web of Science 中的 SCI、SSCI 和 A&HCI 作为数据来源，从国际核心期刊角度出发对科学数据管理研究现状进行了分析，未与国内核心期刊数据进行全面对比；②学术论文作为科学研究成果能够从一定程度上表征科学数据管理主题的研究与发展现状，未能与科学数据管理实践相结合进行多源数据综合分析。这也将是本主题今后进一步研究的方向。

参考文献：

[1] 张先恩. 国家科学数据共享工程[J]. 科学中国人,2004 (9): 11－13.
[2] 国务院. 国务院关于印发促进大数据发展行动纲要的通知[EB/OL]. [2017－12－03]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.

[3] 傅小锋,李俊,黎建辉. 国际科学数据的发展与共享[J]. 中国基础科学,2007(2):30 – 35.
[4] 刘桂锋,卢章平,阮炼. 美国高校图书馆研究数据管理服务内容研究[J]. 图书馆论坛, 2015,35(8):137 – 144.
[5] 司莉,辛娟娟. 英美高校科学数据管理与共享政策的调查分析[J]. 图书馆论坛, 2014, 34(9):80 – 85,65.
[6] 唐燕花. 高校科研数据管理服务实践研究及建议[J]. 图书情报工作,2016,60(24):130 – 138.
[7] 卫军朝,张春芳. 国内外科学数据管理平台比较研究[J]. 图书情报知识,2017(5):97 – 107.
[8] VETTE J, KARLOW N. Data management at National Space Science Data Center[J]. Journal of spacecraft and rockets, 1970,7(10):1234 – 1240.
[9] CAMPBELL W, SMITH P, PRICE R, et al. Advancements in land science data management pilot land data system[J]. Science of the total environment, 1986,56(1): 31 – 44.
[10] GOFFREDO M. The role of scientific data management systems in handling laboratory information[J]. American laboratory, 1999, 31(6): 65 – 69.
[11] FREEDLAND K, CARNEY R. Data management and accountability in behavioral and biomedical-research[J]. American psychologist, 1992,47(5): 640 – 645.
[12] BRUESTLE H, SNIDER D, RUSSELL K, et al. ISDMS - a scientific-data management-system[J]. Software-practice & experience, 1985,15(5): 473 – 487.
[13] 杨行,屈宝强,赫运涛,等. 世界主要国家科学数据资源共享和管理的对比分析和启示[J]. 中国科技资源导刊,2016,(6):18 – 25.
[14] NASA. NASA mission project data management plans (PDMPs) [EB/OL]. [2017 – 12 – 13]. http://nssdc.gsfc.nasa.gov/archive/pdmp/.
[15] NASA. Data & information policy[EB/OL]. [2017 – 12 – 13]. http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/.
[16] JOHNS HOPKINS LIBRARIES. Data management services[EB/OL]. [2017 – 12 – 13]. http://dms.data.jhu.edu/.
[17] WEI J, HE D, YU C. Social science data repositories in data deluge: a case study of ICPSR's workflow and practices[J]. The electronic library, 2017,35(4): 626 – 649.
[18] KIM J. Data sharing from the perspective of faculty in Korea[J]. Libri, 2017,67(3):179 – 192.
[19] DIERKES J, WUTTKE U. The Gottingen Eresearch alliance: a case study of developing and establishing institutional support for research data management [J]. ISPRS international journal of geo-information, 2016,5(8):133.
[20] WITT M. Institutional repositories and research data curation in a distributed environment[J]. Library trends, 2008,52(2):191 – 201.
[21] ADCOCK S, SPIRITO W, GARRETT R. Geochemical data management - issues and solutions [J]. Geochemistry-exploration environment analysis, 2013,13(4): 337 – 348.
[22] FANKHAUSER P, FIEDLER N, WITT A. Research data management in the humanities using the example of germanic linguistics [J]. Zeitschrift fur bibliothekswesen und bibliographie, 2013,60(6):296 – 306.
[23] DOWNS R, CHEN R. Designing submission and workflow services for preserving interdisciplinary scientific data [J]. Earth science informatics, 2010,3(1/2): 101 – 110.

作者贡献说明:
李煜:论文结构设计,内容撰写及后期修改;
刘虹:数据处理,外文资料翻译;
孙建军:研究方向确定,理论指导。

Visualization Analysis of Foreign Scientific Data Management Based on Multi-dimensional Perspectives

Li Yu[1,2] Liu Hong[1] Sun Jianjun[1]
[1] School of Information Management, Nanjing University, Nanjing 210023
[2] Northwest University library, Xi'an 710127

Abstract: [**Purpose/significance**] Under the background of big data, the research of scientific data management has important theoretical and practical significance. [**Method/process**] Based on WoS and Google Scholar databases, the paper utilized Citespace, Ucinet, TDA and other analysis tools to explore the evolution of scientific data management research in international academic community. [**Result/conclusion**] The paper fully revealed the history, current situation and development direction of scientific data management research from four aspects including time distribution map, spatial distribution map, thematic evolution map and knowledge base map. Furthermore, the paper put forward relevant proposals according to the domestic research status quo.

Keywords: scientific data data management knowledge mapping visualization analysis